

Weekly report

1 Done.

1.1 Vis Presentation

Improve the power point based on the advices given by Jia-Kai.

Practice for fast forward and presentation.

Prepare for questions:

- Q: More detail about delta view?
A: We use red and green to represent the difference in the distribution. The opacity represents the absolute value of the differences.
- Q: How to change granularity?
A: We suppose that records are evenly distributed in each square, then we can count the amount based on other granularity.
- Q: How many dimensions does your system can deal with?
A: I guess you mean how many necessary attributes can we deal with. According to the limitation of the screen, I guess it is about seven. As for the group number, you must know, the more accurate the information, the easier it is to expose the user's identity, so sometimes it is meaningless.
- Q: Why do you employ both prop-nodes and attr-nodes?
A: Some privacy issues are produced by one attribute division, while others may be produced by joint division. The prop-node is related to the joint one. Besides, there is a switch on the interface, which can control the display of prop-nodes.
- Q: What is your next direction?
A: Actually, this is a new field for visualization. There are a lot of work waiting for everyone to complete, for example, high-dimensional data processing, different data formats and so on.
- Q: Why you express tabular data by tree?
A: First, if we order the attributes in a good way, the PER-Tree will have a good behavior on finding the source of the privacy exposure risk. If users observe all attributes together at the first beginning, they will be confused, like what should I do? Second, the pruning features of tree is really useful when you analyze partial data.
- Q: Why you employ three kinds of syntactic models? Whether the t-closeness can replace the other two?
A: Actually not. One of the drawbacks of t-closeness is that there is no standard way to calculate the distance of distribution. Besides, if what we want to observe is the difference between two kinds of individuals, we can't reduce the difference of distribution. There will be same situation for l-diversity.
- Q: What is the system better than the automatic model?

A: As I mentioned here, visualization has its own advantages. For example, with our system, you can customize your own privacy needs in the same dataset. For example, we need protect children better than adults.

- Q: In UPD-Matrix, the complex expressions can cause confusion.

A: Yes, that is a good question. However, we want to provide information as specific as possible. Maybe this view is not so intuitive, so we provide quantitative metric and delta view.

- Q: More detail about quantitative metric?

A: It is not about the entire distribution. It is the attribute value change of each record. (We use the group values instead of the data values that were processed.) We think it is more useful to study the multi-dimensional relationship, which can't be observed in the UPD-Matrix.

1.2 Geo-privacy

We add detailed instruction to the interface, so that participants won't make mistakes.



Besides, when participants switch to next task, there will be a hint like bellow.



1.3 Graph privacy

We reviewed the entire idea together. To simplify the algorithms, we decided to solve all privacy issues by adding edge. Besides, we will focus on locate the issues and solve them effectively, so we won't take combination optimization into consideration.

2 Progress

| Item | Deadline | Current progress | Remark |
|---------------|----------|--------------------|--------|
| Geo-privacy | EuroVis | Add introductions. | - |
| Graph privacy | ?? | Make plan. | - |